Adaptive Multi-Agent Systems for Constrained Optimization

William Macready[1,2], **Stefan Bieniawski**[3], and **David H. Wolpert**[1,4]
1 - MS 269-1, NASA Ames Research Center, Moffett Field, CA, 94035
2 - wgm@ptolemy.arc.nasa.gov
3 - 253 Durand, Dept. of Aeronautics, Stanford, CA 94305, stefanb@stanford.edu
4 - dhw@ptolemy.arc.nasa.gov

Abstract

Product Distribution (PD) theory is a new framework for analyzing and controlling distributed systems. Here we demonstrate its use for distributed stochastic optimization. First we review one motivation of PD theory, as the information-theoretic extension of conventional full-rationality game theory to the case of bounded rational agents. In this extension the equilibrium of the game is the optimizer of a Lagrangian of the (probability distribution of) the joint state of the agents. When the game in question is a team game with constraints, that equilibrium optimizes the expected value of the team game utility, subject to those constraints. The updating of the Lagrange parameters in the Lagrangian can be viewed as a form of automated annealing, that focuses the MAS more and more on the optimal pure strategy. This provides a simple way to map the solution of any constrained optimization problem onto the equilibrium of a Multi-Agent System (MAS). We present computer experiments involving both the Queen's problem and K-SAT validating the predictions of PD theory and its use for off-the-shelf distributed adaptive optimization.

Introduction

Product Distribution (PD) theory was introduced recently in (Wolpert 2003; 2004b; 2004a). It is a broad framework for analyzing, controlling, and optimizing distributed systems. Among its potential applications are (constrained) optimization, distributed adaptive control of multi-agent systems, sampling of probability densities, density estimation, numerical integration, reinforcement learning, information-theoretic bounded rational game theory, population biology, and management theory. See (Antoine *et al.* 2004; Airiau & Wolpert 2004; Lee & Wolpert 2004; Bieniawski & Wolpert 2004).

Here we demonstrate its use for distributed stochastic optimization. Typically in stochastic approaches to optimization one uses probability distributions to help search for point $x \in X$ optimizing a function $G(x)$. In contrast, in the PD approach one searches for a probability distribution $P(x)$ that optimizes an associated Lagrangian, $\mathcal{L}_G(P)$. Since P is a vector in a Euclidean space, the search can be done via techniques like gradient descent or Newton's method — even if X is a categorical, finite space.

One motivation of PD theory is as the information-theoretic extension of conventional full-rationality game theory to the case of bounded rational agents. Information theory shows that the equilibrium of a game played by bounded rational players is the optimizer of a Lagrangian of the probability distribution of the agents' joint-moves. From this perspective distributed adaptive optimization algorithms in which each agent uses reinforcement learning are just one — inefficient — way to optimize $\mathcal{L}$ (Wolpert 2003; 2004b).

In any game, bounded rational or otherwise, the agents are independent, with each agent i choosing its move x_i at any instant by sampling its probability distribution (mixed strategy) at that instant, $q_i(x_i)$. Accordingly, the distribution of the joint-moves is a product distribution, $P(x) = \prod_i q_i(x_i)$. In this representation of a Multi-Agent System (MAS), all coupling between the agents occurs indirectly; it is the separate distributions of the agents $\{q_i\}$ that are statistically coupled, while the actual moves of the agents are independent. This is reflected in the fact that the optimization of the Lagrangian (e.g., via gradient descent) can be done in a completely distributed manner.

When the game in question is a team game with constraints, the bounded rational equilibrium optimizes the expected value of the team game utility, subject to those constraints and an overall entropy value. Updating of the Lagrange parameters in the usual way provides a form of automated annealing, focusing the MAS more and more on the optimal pure strategy as the parameters change. This provides a broadly applicable way to cast any constrained optimization problem as the equilibrating process of a MAS, together with an efficient method for that equilibrating process.

In the next section we review the game-theory motivation of PD theory. We then present details of our Lagrangian-minimization algorithm. We end with computer experiments involving both the N Queen's problem and K-sat (Yokoo & Hirayama 2000; Mezard, Parisi, & Zecchina 2002). These results, though preliminary, validate the predictions of PD theory, and indicate its usefulness as a general purpose technique for distributed solution of constrained optimization problems.

Bounded Rational Game Theory

Review of noncooperative game theory

In noncooperative game theory one has a set of N **players**. Each player i has its own set of allowed **pure strategies**. A **mixed strategy** is a distribution $q_i(x_i)$ over player i's possible pure strategies. Each player i also has a **private utility** function g_i that maps the pure strategies adopted by all N of the players into the real numbers. So given mixed strategies of all the players, the expected utility of player i is $E(g_i) = \int dx \ \prod_j q_j(x_j) g_i(x)$ [1].

In a **Nash equilibrium** every player adopts the mixed strategy that maximizes its expected utility, given the mixed strategies of the other players. More formally, $\forall i, q_i = \text{argmax}_{q'_i} \int dx \ q'_i \prod_{j \neq i} q_j(x_j) \ g_i(x)$. Perhaps the major objection that has been raised to the Nash equilibrium concept is its assumption of **full rationality** (Fudenberg & Levine 1998; Fudenberg & Tirole 1991). This is the assumption that every player i can both calculate what the strategies $q_{j \neq i}$ will be and then calculate its associated optimal distribution. In other words, it is the assumption that every player will calculate the entire joint distribution $q(x) = \prod_j q_j(x_j)$. If for no other reasons than computational limitations of real humans, this assumption is essentially untenable.

Review of the maximum entropy principle

Shannon was the first person to realize that based on any of several separate sets of very simple desiderata, there is a unique real-valued quantification of the amount of syntactic information in a distribution $P(y)$. He showed that this amount of information is (the negative of) the Shannon entropy of that distribution, $S(P) = -\int dy \ P(y) \ln[\frac{P(y)}{\mu(y)}]$. So for example, the distribution with minimal information is the one that doesn't distinguish at all between the various y, i.e., the uniform distribution. Conversely, the most informative distribution is the one that specifies a single possible y. Note that for a product distribution, entropy is additive, i.e., $S(\prod_i q_i(y_i)) = \sum_i S(q_i)$.

Say we are given some incomplete prior knowledge about a distribution $P(y)$. How should one estimate $P(y)$ based on that prior knowledge? Shannon's result tells us how to do that in the most conservative way: have your estimate of $P(y)$ contain the minimal amount of extra information beyond that already contained in the prior knowledge about $P(y)$. Intuitively, this can be viewed as a version of Occam's razor. This approach is called the maximum entropy (maxent) principle. It has proven useful in domains ranging from signal processing to supervised learning (Mackay 2003).

Maxent Lagrangians

Much of the work on equilibrium concepts in game theory adopts the perspective of an external observer of a game. We are told something concerning the game, e.g., its utility functions, information sets, etc., and from that wish to predict what joint strategy will be followed by real-world players of the game. Say that in addition to such information, we are told the expected utilities of the players. What is our best estimate of the distribution q that generated those expected utility values? By the maxent principle, it is the distribution with maximal entropy, subject to those expectation values.

To formalize this, for simplicity assume a finite number of players and of possible strategies for each player. To agree with the convention in other fields, from now on we implicitly flip the sign of each g_i so that the associated player i wants to minimize that function rather than maximize it. Intuitively, this flipped $g_i(x)$ is the "cost" to player i when the joint-strategy is x, though we will still use the term "utility".

Then for prior knowledge that the expected utilities of the players are given by the set of values $\{\epsilon_i\}$, the maxent estimate of the associated q is given by the minimizer of the Lagrangian

$$\mathcal{L}(q) \equiv \sum_i \beta_i [E_q(g_i) - \epsilon_i] - S(q) \quad (1)$$

$$= \sum_i \beta_i \Big[\int dx \prod_j q_j(x_j) g_i(x) - \epsilon_i \Big] - S(q) \quad (2)$$

where the subscript on the expectation value indicates that it evaluated under distribution q, and the $\{\beta_i\}$ are "inverse temperatures" $\beta_i = 1/T_i$ implicitly set by the constraints on the expected utilities.

Solving, we find that the mixed strategies minimizing the Lagrangian are related to each other via

$$q_i(x_i) \propto e^{-E_{q_{(i)}}[G|x_i]} \quad (3)$$

where the overall proportionality constant for each i is set by normalization, and $G(x) \equiv \sum_i \beta_i g_i(x)$.[2] In Eq. (3) the probability of player i choosing pure strategy x_i depends on the effect of that choice on the utilities of the other players. This reflects the fact that our prior knowledge concerns all the players equally.

If we wish to focus only on the behavior of player i, it is appropriate to modify our prior knowledge. To see how to do this, first consider the case of maximal prior knowledge, in which we know the actual joint-strategy of the players, and therefore all of their expected costs. For this case, trivially, the maxent principle says we should "estimate" q as that joint-strategy (it being the q with maximal entropy that is consistent with our prior

[1]Throughout this paper, the integral sign is implicitly interpreted as appropriate, e.g., as Lebesgue integrals, point-sums, etc.

[2]The subscript $q_{(i)}$ on the expectation value indicates that it is evaluated according the distribution $\prod_{j \neq i} q_j$. The expectation is conditioned on player i making move x_i.

knowledge). The same conclusion holds if our prior knowledge also includes the expected cost of player i.

Modify this maximal set of prior knowledge by removing from it specification of player i's strategy. So our prior knowledge is the mixed strategies of all players other than i, together with player i's expected cost. We can incorporate prior knowledge of the other players' mixed strategies directly, without introducing Lagrange parameters. The resultant **maxent Lagrangian** is

$$\begin{aligned}\mathcal{L}_i(q_i) &\equiv \beta_i[\epsilon_i - E(g_i)] - S_i(q_i) \\ &= \beta_i[\epsilon_i - \int dx \prod_j q_j(x_j) g_i(x)] - S_i(q_i)\end{aligned}$$

solved by a set of coupled **Boltzmann distributions**:

$$q_i(x_i) \propto e^{-\beta_i E_{q_{(i)}}[g_i|x_i]}. \tag{4}$$

Following Nash, we can use Brouwer's fixed point theorem to establish that for any non-negative values $\{\beta\}$, there must exist at least one product distribution given by the product of these Boltzmann distributions (one term in the product for each i).

The first term in $\mathcal{L}_i$ is minimized by a perfectly rational player. The second term is minimized by a perfectly *irrational* player, i.e., by a perfectly uniform mixed strategy q_i. So β_i in the maxent Lagrangian explicitly specifies the balance between the rational and irrational behavior of the player. In particular, for $\beta \to \infty$, by minimizing the Lagrangians we recover the Nash equilibria of the game. More formally, in that limit the set of q that simultaneously minimize the Lagrangians is the same as the set of delta functions about the Nash equilibria of the game. The same is true for Eq. (3).

Eq. (3) is just a special case of Eq. (4), where all player's share the same private utility, G. (Such games are known as **team games**.) This relationship reflects the fact that for this case, the difference between the maxent Lagrangian and the one in Eq. (2) is independent of q_i. Due to this relationship, our guarantee of the existence of a solution to the set of maxent Lagrangians implies the existence of a solution of the form Eq. (3). Typically players will be closer to minimizing their expected cost than maximizing it. For prior knowledge consistent with such a case, the β_i are all non-negative.

For each player i define

$$f_i(x, q_i(x_i)) \equiv \beta_i g_i(x) + \ln[q_i(x_i)].$$

Then the maxent Lagrangian for player i is

$$\mathcal{L}_i(q) = \int dx\ q(x) f_i(x, q_i(x_i)). \tag{6}$$

Now in a bounded rational game every player sets its strategy to minimize its Lagrangian, given the strategies of the other players. In light of Eq. (6), this means that we interpret each player in a bounded rational game as being perfectly rational for a utility that incorporates its computational cost. To do so we simply need to expand the domain of "cost functions" to include probability values as well as joint moves.

Often our prior knowledge will not consist of exact specification of the expected costs of the players, even if that knowledge arises from watching the players make their moves. Such alternative kinds of prior knowledge are addressed in (Wolpert 2004b; 2004a). Those references also demonstrate the extension of the formulation to allow multiple utility functions of the players, and even variable numbers of players. Also discussed there are **semi-coordinate** transformations, under which, intuitively, the moves of the agents are modified to set in binding contracts.

Optimizing the Lagrangian

Given that the agents in a MAS are bounded rational, if we have them play a constrained team game with world utility G, their equilibrium will be the optimizer of G subject to those (potentially inexact) constraints (Wolpert 2003; 2004a; Bieniawski & Wolpert 2004). Formally, let $\{c_j(x)\}$ be the constraint functions, i.e., we seek a joint-move x such that all of the $\{c_j(x)\}$ are nowhere negative. Then the bounded rational equilibrium will minimize the Lagrangian, Eq. (2), where the world world utility is augmented with Lagrange multipliers, λ_j, for each of the

$$G(x) \to G(x) + \textstyle\sum_j \lambda_j c_j(x) \tag{7}$$

Consider a fixed set of values for the Lagrange parameters. We can minimize the associated Lagrangian using gradient descent, since the gradient can be evaluated in closed form. We can also evaluate the Hessian in closed form. This allows us to use **constrained Newton's method**. This is a variant of Newton's method in which we first modify the Lagrangian, and then enforce both independence of the agents, and that the search stays on the simplex of valid probabilities (Wolpert 2004a; Antoine *et al.* 2004; Bieniawski & Wolpert 2004):

$$\begin{aligned}q_i(x_i) \to\ & q_i(x_i) - \alpha q_i(x_i) \times \\ & \left\{ \frac{E[G|x_i] - E[G]}{T} + S(q_i) + \ln q_i(x_i) \right\}\end{aligned} \tag{8}$$

where α plays the role of a step size. The Lagrange multipliers are then updated in the usual way, by taking the partial derivatives of the augmented Lagrangian:

$$\lambda_j \to \lambda_j + \eta E[c_j(x)] \tag{9}$$

where η is a separate step size.

The update of Eq. (8) involves a separate conditional expected utility for each agent. These are estimated either exactly if a closed form expression is available or with Monte-Carlo sampling if no simple closed form exists. In Monte Carlo sampling the agents repeatedly and jointly IID sample their probability distributions to generate joint moves, and the associated utility values are recorded. Since accurate estimates usually requires extensive sampling, we replace the G occurring in each agent i's update rule with a private utility g_i chosen to ensure that the Monte Carlo estimation of $E(g_i|x_i)$ has both low bias (with respect to estimating $E(G|x_i)$ and low variance (Duda, Hart, & Stork 2000). Intuitively, this bias reflects the alignment between the private and world utilities. At zero bias, reducing private utility necessarily reduces world utility. Variance instead reflects how much the utility depends on the agent's own move rather than those of the other agents. With low variance, the agents can perform the individual optimizations accurately with minimal Monte-Carlo sampling.

In this paper we concentrated on two types of private utility in addition to the team game (TG) utility. The first is the **Aristocrat Utility** (AU) utility. It is the utility, out of all those guaranteed to have zero bias, that has minimal variance:

$$g_{AU_i}(x_i, x_{(i)}) = G(x_i, x_{(i)}) - \sum_{x'_i} \frac{N^{-1}_{x'_i}}{\sum_{x''_i} N^{-1}_{x''_i}} G(x'_i, x_{(i)}) \tag{10}$$

where N_{x_i} is the number of times that agent i makes move x_i in the most recent set of Monte Carlo samples. In addition we consider the **Wonderful Life Utility** (WLU), which is an approximation to AU that also has zero bias:

$$g_{WLU_i}(x_i, x_{(i)}) = G(x_i, x_{(i)}) - G(CL_i, x_{(i)}) \tag{11}$$

where the clamping value CL_i fixes agent i's move to the one to which it assigns lowest probability action (Wolpert 2003; 2004a).

Experiments

N-Queens Problem

The N-queens problem is not hard to solve, especially with centralized algorithms (Sosič & Gu 1990). However it is a good illustration and testbed of the PD-theory approach. The goal in this problem is to locate N queens on a $N \times N$ chessboard such that there are no attacks between any of the queens, i.e., no shared rows, columns or diagonals. For the results presented here, $N = 8$ and each agent's move is the position of a queen on an associated row of the chessboard. Denoting agent i's making move j as $x_i(j)$, the constraints are

$$\begin{aligned} x_i(j) &\neq x_k(j) && (12)\\ x_i(j) &\neq x_{i+k}(j+k) \neq x_{i-k}(j-k) && (13)\\ x_i(j) &\neq x_{i+k}(j-k) \neq x_{i-k}(j+k) && (14) \end{aligned}$$

For 8 queens this results in 84 constraints.

For this study the step size α was set to 1.0, while the data aging rate γ was set to 0.5. Data aging is used to weight of the previous samples relative to the most recent ones and smoothes the effect of the Monte-Carlo sampling. The optimizations were performed at a range of fixed temperatures and 10 Monte-Carlo samples were used for each probability and Lagrange multiplier update. We concentrated on the number of iterations to convergence, i.e., the number of probability updates times the number of Monte-Carlo samples per update, for 50 random trials of the problem. The optimization was terminated when a single Monte-Carlo sample within an iteration was found which satisfied all of the constraints. Figure 1 shows the effect of varying the agent utilities. The first plot shows the completion rate, i.e., the percentage of the trials in which a solution was found in under 10000 iterations. The second plot shows the mean samples for those cases which did find a solution while the third plot shows the 95% confidence level in this mean.

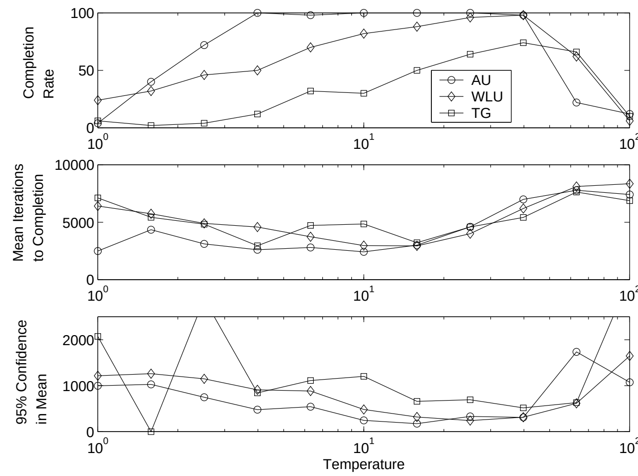

Figure 1: Performance versus temperature.

The results clearly indicate the advantage of using AU, which resulted in high completion rate over a wide range of temperatures. In addition the mean samples to completion was low over a wide range, with a minimum of 2500. WLU also performed quite well, although the high completion rate was only over a limited temperature range and was accompanied by an increase in the iterations to convergence.

Figure 2 provides a more detailed look at the distribution of iterations to convergence. The curves plot the cumulative probability for the various utilities at the temperature giving the highest completion rate for that utility. Again the benefit of AU is clear in ensuring that all of the cases reach convergence. Note that for AU, and somewhat for WLU, the tail of the distribution is short, whereas for TG it is elongated. Generally changing the temperature increased the completion rate and the mean iterations to completion but gave tighter distributions. A sudden drop in completion rate at high temperature is due to the mean completion rate approaching the maximum allowed number of iterations.

K-sat

In the K-sat problem there are N binary variables $x_i \in \{0, 1\}$ and C clauses. The i'th such clause involves K variables labelled by $v_{i,j}$ (for $j \in \{1, \cdots, K\}$), and K binary values associated with each i and labelled by $\sigma_{i,j}$. The ith clause is satisfied iff $c_i(x) \equiv$

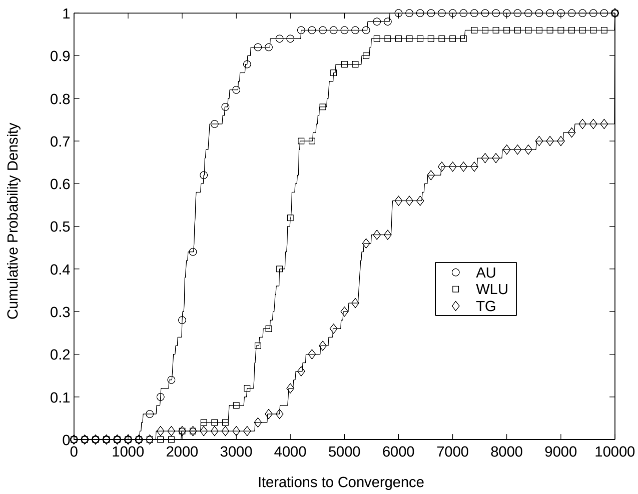

Figure 2: Cumulative probability at best temperature.

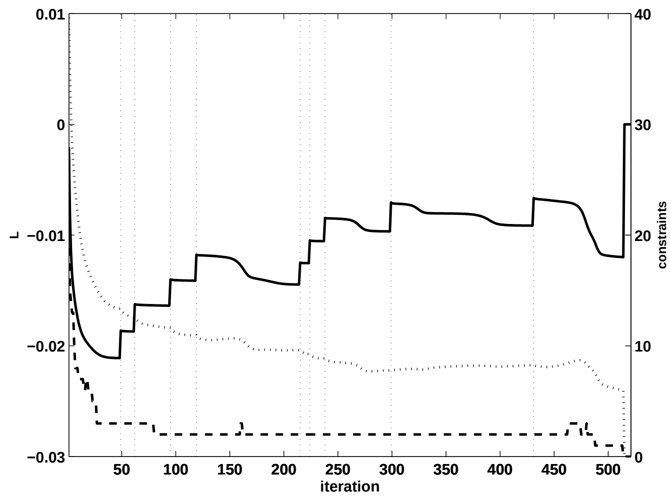

Figure 3: Evolution of Lagrangian value (solid line), expected constraint violation (dotted line), and constraint violations of most likely configuration (dashed line).

$\bigvee_{j=1}^{K}[x_{v_{i,j}} = \sigma_{i,j}]$ is true. Accordingly we write $G(x,\lambda) \equiv \sum_i \lambda_i \bigvee_{j=1}^{K}[x_{v_{i,j}} = \sigma_{i,j}] = \lambda^\top c(x)$ where λ and c are vectors of length C whose i components are λ_i, and $c_i(x)$ respectively.

Unlike the queens example, here we illustrate PD theory using exact evaluation of the required expectations rather than Monte Carlo sampling. Noting that the ith clause is violated only when all $x_{v_{i,j}} = \overline{\sigma}_{i,j}$ (with $\overline{\sigma} \equiv \text{not}\,\sigma$), the Lagrangian over product distributions can be written as

$$\mathcal{L}(q) = \sum_{i=1}^{C}\Big\{\lambda_i \prod_{j=1}^{K} q_{v_{i,j}}(\overline{\sigma}_{i,j}) - TS(q_i)\Big\} = \lambda^\top c(q) - TS(q). \tag{15}$$

where $c(q)$ is the C-vector of expected constraint violations whose ith component is $c_i(q) \equiv \prod_{j=1}^{K} q_{v_{i,j}}(\overline{\sigma}_{i,j})$, and where $S(q) \equiv \sum_i S(q_i)$ is the usual entropy function. We assign a player to each binary variable, so the only communication required to evaluate G and its appropriate derivatives is between agents appearing in the same clause. Typically then, the communication network is sparse. For the $N = 100$, $K = 3$, $C = 430$ variable problem we address here each agent interacts with only 6 other agents on average.

For any fixed setting of the Lagrange multipliers, the Lagrangian is minimized according to the constrained Newton update given in Eq. (8). When the constrained Newton step would move out of the feasible region (i.e. having $q_i(\sigma) < 0$ or $q_i(\sigma) > 1$) we move to that point in the feasible region nearest to the proposed update. The minimization is terminated at a local minimum q^* of the Lagrangian which is recognized when the Newton step becomes sufficiently small. If all constraints are satisfied at q^* we terminate and return the solution $x^* = \operatorname{argmax}_x q(x)$. If not all constraints are satisfied the Lagrange multipliers are updated in the standard manner according to the constraint violation, i.e. $\lambda_i \to \lambda_i + \eta c_i(q^*)$ where η is some step size. In the present context, this standard updating rule for constrained minimization offers a number of appealing benefits. Firstly, those constraints which are violated most strongly have their penalty increased the most, and consequently, the agents involved in those constraints are most likely to alter their state. Secondly, the Lagrange multipliers contain a history of the constraint violations (since we keep adding to λ) so that when the agents coordinate on their next move they are unlikely to return a previously violated state since those configurations will also have higher than average Lagrange multiplier values. This mimics the approach used in taboo search where revisiting of configurations is explicitly prevented, and aids in an efficient exploration of the search space. Lastly, we note that rescaling the Lagrangian by the norm of λ gives $\mathcal{L}(q) = \hat{\lambda}^\top c(q) - TS(q)/\|\lambda\|$ where $\hat{\lambda} = \lambda/\|\lambda\|$ so that the updating the Lagrange multipliers can be seen as defining a cooling schedule where $T \to T/\|\lambda\|$. The parameter η thus governs the overall rate of cooling. We used $\eta = 0.5$ in our experiments.

We present results for a 100 variable $K = 3$ problem. The problem is a satisfiable formula called `uf100-01.cnf` available from SATLIB at `www.satlib.org`. It was generated with the ratio of clauses to variables being near the phase transition, and is consequently a difficult problem. Figure 3 shows the variation of the Lagrangian, the expected number of constraint violations $1^\top c(q)$, and the number of constraints violated in the most probable state $x_{\text{mp}} \equiv \operatorname{argmax}_{x'} q(x')$ as a function of the number of iterations. The starting state is the maximum entropy configuration having all $q_i = [1/2\ 1/2]$, and the starting temperature is $T = 0.0015$. The iterations at which the Lagrange multipliers are updated are indicated by vertical

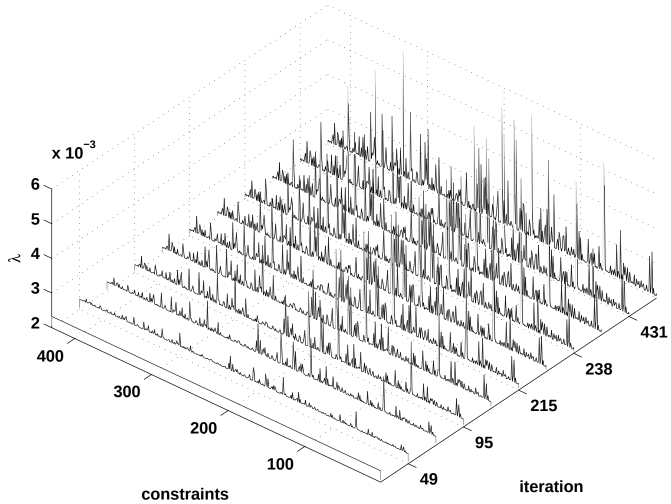

Figure 4: Each constraint's Lagrange multiplier versus the iterations when they change.

dashed lines, and are clearly visible as discontinuities in the Lagrangian values.

The temperature is immediately dropped to 0 when $x_{\rm mp}$ satisfies all constraints. Figure 4 shows the evolution of the Lagrange multipliers. Initially, λ is set to 1 and as time progresses the history of the multipliers is clearly evident. We also present a figure showing the evolution of the probability density of constraint violations. This exact approach where we evaluate the expected value of $\lambda^\top c$ exactly obscures the probabilistic nature of the search. To show the stochastic underpinnings of the algorithm we plot in Figure 5 the probability density of $G = 1^\top c$ (i.e. the number of constraint violations) obtained as $\text{Prob}(G) = \sum_x q(x)\delta\big(G - G(x,1)\big)$. In determining the density 10^4 samples we drawn from $q(x)$ with Gaussians centered at each value of $G(x, \mathbf{1})$ and with the width of all Gaussians determined by cross validation of the log likelihood. The fact that there is non-zero probability of obtaining non-integral numbers of constraint violations is an artifact of the finite width of the Gaussians.

Conclusion

A distributed constrained optimization framework based upon product distribution theory has been presented. Motivation for the framework was drawn from the extension of full-rationality game theory to the case of bounded rational agents. An algorithm was developed and demonstrated on two distributed constraint satisfaction problems, the N-queen's problem and K-Sat. The results show a promising approach for highly distributed constrained optimization.

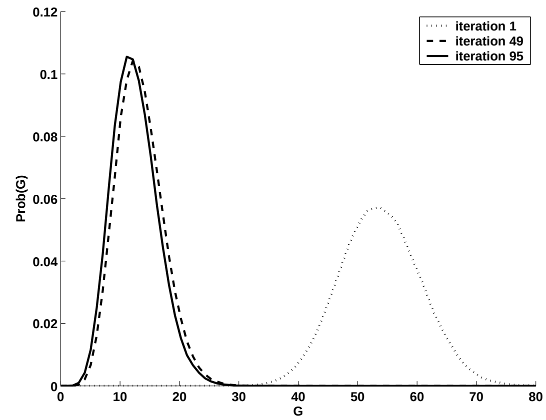

Figure 5: $P(G)$ after minimizing the Lagrangian for the first 3 multiplier settings. At termination $P(G) = \delta(G)$.

References

Airiau, S., and Wolpert, D. H. 2004. Product distribution theory and semi-coordinate transformations. Submitted to AAMAS 04.

Antoine, N.; Bieniawski, S.; Kroo, I.; and Wolpert, D. H. 2004. Fleet assignment using collective intelligence. In *Proceedings of 42nd Aerospace Sciences Meeting*. AIAA-2004-0622.

Bieniawski, S., and Wolpert, D. H. 2004. Adaptive, distributed control of constrained multi-agent systems. Submitted to AAMAS 04.

Duda, R. O.; Hart, P. E.; and Stork, D. G. 2000. *Pattern Classification (2nd ed.)*. Wiley and Sons.

Fudenberg, D., and Levine, D. K. 1998. *The Theory of Learning in Games*. Cambridge, MA: MIT Press.

Fudenberg, D., and Tirole, J. 1991. *Game Theory*. Cambridge, MA: MIT Press.

Lee, C. F., and Wolpert, D. H. 2004. Product distribution theory for control of multi-agent systems. Submitted to AAMAS 04.

Mackay, D. 2003. *Information theory, inference, and learning algorithms*. Cambridge University Press.

Mezard, M.; Parisi, G.; and Zecchina, R. 2002. Analytic and algorithmic solution of random satisfiability problems. *Science*.

Sosič, R., and Gu, J. 1990. A polynomial time algorithm for the N-queens problem. *SIGART Newsletter (Special Interest Group on Artificial Intelligence)* 1.

Wolpert, D. H. 2003. Factoring a canonical ensemble. cond-mat/0307630.

Wolpert, D. H. 2004a. Bounded rationality game theory and information theory. Submitted.

Wolpert, D. H. 2004b. Information theory — the bridge connecting bounded rational game theory and statistical physics. In D. Braha, A. M., and Bar-Yam, Y., eds., *Complex Engineering Systems*.

Yokoo, M., and Hirayama, K. 2000. Algorithms for distributed constraint satisfaction: A review. *Autonomous Agents and Multi-Agent Systems* 3:198–212.